# Morse Code Audio Recognition using LSTM-CTC Model

Arkar Win, Kyawt Kyawt San
University of Information Technology,
Yangon, Myanmar
E-mail: arkarwin@uit.edu.mm,
kyawtkyawtsan@uit.edu.mm

*Abstract*— **In media transmission, Morse code is utilized to encode text characters as sequences of two different duration of a signal. Morse code audio is also utilized for SOS of maritime. The popular audio recognition algorithms are Hidden Markov Model, Gaussian Mixture Model, LSTM, and so on. This paper focuses on Morse code audio recognition using LSTM. Long Short-Term Memory (LSTM) is selected because it is the best deep learning algorithm for the sequence-to-sequence model. Morse code audio dataset was created text characters to Morse code audio. Mel-frequency Cepstral Coefficients (MFCC) algorithm is used to extract features of the dataset. LSTM handles to get related information of each Morse code audio by using MFCCs features. Connectionist Temporal Classification (CTC) decides to get the related label by using LSTM information. This paper aims to help the Morse code audio listener by getting the right information about the input data.**

***Keywords— Morse code, MFCC, LSTM, CTC, Audio recognition, Feature extraction***

## I. INTRODUCTION

Morse code is an alphabet character or code in which or sound signals. It can be created with a signal of some sort, whether that be a written or symbolic image, flashing a light. It is commonly used in radio communication, communication for soldiers in war. Morse code was critical for communication during World War II. It is the standard format for ocean communication. It is also a communication tool for physically disabled persons who are suffered from a severe handicap, muscle atrophy, and cerebral palsy. The advantages of Morse are simple coding methods, strong noisy immunity, ease of implementation.

Morse code can be represented with the English letters, numbers, and punctuation marks in a different order. The code contains five types. They are dot, dash, space between them within the character, space between characters, and word space. The proportion of dot to dash is about 1:3. The silent ratio (space between dot, dash within character: space between characters: word space) is about 1:3:5 depending on the structure of the Morse code.

Morse code have many problems to get information. The problems are learning the Morse code, time-consuming, interrupting. Recognition can help the problems. Example, the disabled person gives the information what he wants to say by using the morse code generator machine. The information of disabled person is coming with morse code audio and that is needed to convert as readable format. The communication with encoded data is very useful in some places. The only problem is the encoded data are needed to convert readable information for receivers. The system aims to help getting right readable information from Morse code audios. Normally, morse signal with the high signal is not difficult to recognize. However, when the background noise is high and the signal-to-noise is very low, the traditional way is difficult to signal recognition. This paper uses the MFCC algorithm to get audio features from the raw signals. Then, the LSTM algorithm handles getting related information by using the training process. CTC decides to get the related label of the raw signal by using LSTM information.

The remainder of this paper is sorted out as follows. The related work is introduced in Section 2. The background theory for Morse code recognition is introduced in Section 3. Section 4 focuses on the proposed system architecture of Morse code audio recognition. The experimental result is described in Section 5. Section 6 includes the conclusion and future work.

## II. RELATED WORK

There are various algorithms for speech recognition. But a few papers take research for Morse code audio recognition. The related work of Morse code audio recognition is below.

Xianyu Wang [1] presented automatic Morse code recognition using MFCC and recognition algorithm. It proposed the new MFCC architecture to recognize the Morse code audio to get the better result. The new MFCC architecture use a linear filter bank instead of using a normal Mel-filter bank. And the experimental result compared the normal feature extraction and new feature extraction result.

Ming Liu [2] proposed the model to enhance the speech signal by using MFCC and LSTM. It used MFCC to enhance wave features and LSTM generates the enhanced speech of input speech. And it used four LSTM layers in the model.

De Mauri [3] described Morse code audio recognition using CNN-LSTM-CTC model. Its model contained 5 CNN layers, 2 LSTM layers, CTC loss, and a decoding layer. The morse code audio passes the CNN layers as feature extraction and recognize the label in LSTM-CTC layers. The result showed on the signal to noise ratio of audio files.

Yang [4] presented the Morse code recognition using a neural network. It processed the Morse code digital steam as the data input. It split the recognition into tone recognition, space recognition, and character recognition. It used the artificial neural network to get the dot or dash result.

Ericks Rachmat Swedia [5] presented the speech recognition using the LSTM network with MFCC and LPC. It built the model with MFCC or LPC, LSTM layers, fully connected layers, SoftMax layer. The result compared the MFCC feature extraction and LPC feature extraction on the model.

Paribesh Regmi [6] presented the Nepali Speech recognition with the RNN-CNN model. The model contains the input layer, LSTM layer, Dense layer, SoftMax layer, and CTC layer. LSTM layer generates the long-term sequence patterns of input audio. Nepali text is trained in the CTC layer.

And CTC layer decides the label of audio. The model illustrated the Nepali audio to convert Nepali text format.

The difference of related work with the propose system are the following statements. The propose system is using new dataset. The system uses MFCC feature extraction instead of CNN. And, the system can directly recognize the alphabet words from Morse code audios.

## III. BACKGROUND THEORY

### A. *Long Short-Term Memory (LSTM)*

Long Short-Term Memory (LSTM) is a modified theory from Recurrent Neural Network (RNN). It can process on a single data point (image) or sequences of data (audio, video). LSTMs network has a chain-like structure with sequence connected LSTM cells. LSTM cell of the working structure contains 6 kinds of steps. LSTM cell that can pass information through the LSTM network. The data of the past time steps can measure into next time ventures by decreasing the impacts of transient memory. As the cell state goes on its handling, data is added or taken out to the cell state with gates. The gates can be prepared what data is proper to keep or overlook during the training process. The gates make a decision which information can keep or pass on the cell state. The gates are - forget gate, input gate, and output gate. The forget gate chooses what data ought to be discarded or kept. The input gate decides in which information is updated by transforming the values. The output gate makes a decision to carry information that should be processed as input of next cell state. The LSTM cell process as

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{1}$$

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{2}$$

$$\tilde{C}_t = tanh(W_c.[h_{t-1}, x_t] + b_c) \tag{3}$$

$$C_t = (f_t * C_{t-1} + i_t * \tilde{C}_t) \tag{4}$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * tanh\,(C_t) \tag{6}$$

Where,

$i_t$ = Input/Update Gate Activation of t state

$f_t$ = Forget Gate Activation of t state

$\tilde{C}_t$ = Candid Memory Cell value of t state

$C_t$ = New Memory Cell value of t state

$o_t$ = Final Output Gate Activation of t state

$h_t$ = Final Output Gate value of t state

$b_i$ = Bias value of Input/Update Gate

$b_f$ = Bias value of Forget Gate

$b_c$ = Bias value of Memory Cell

$b_o$ = Bias value of Output Gate

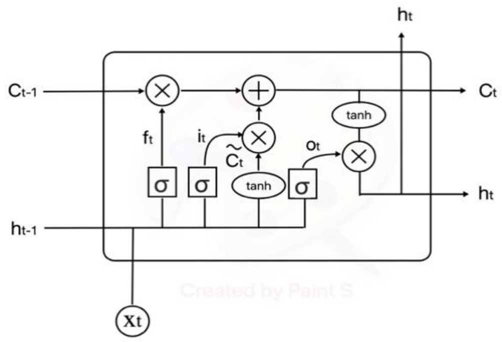


Fig. 1. LSTM cell basic structure

In Figure 1[16], final output gate value of previous cell (h of t-1) and current cell of input (x of t) are used as the input of the cell state. The forget gate (f of t) get the information from the previous hidden state and information from the current input is passed through the sigmoid function. The input and update gate will generate value that is multiply the sigmoid output of current input value (i of t) with the tanh output of current input value. In cell state of that figure, new memory cell value can be gotten by combination of the input and update gate value and multiplied value of previous cell value with forget gate value. The current output gate value can be gotten by multiplication of tanh output of current cell value with sigmoid output of input of the cell state (o of t).

LSTM network is used for classifying, making predictions for time series data. Different types of LSTM networks are as same as the Recurrent Neural Network. They are one to one (is used for image classification), one to many (is used for image caption with sequence word output), many to one (is used for sentence classification), many to many (is used for converting from one language to another language), bidirectional many to many (is used for video classification as the label of each frame).

### B. *Mel-frequency Cepstral Coefficients*

Mel-frequency cepstrum (MFC) can be represented as the power spectrum of a signal or sound. Mel-frequency cepstral coefficients (MFCCs) are coefficients that can build with a combination of MFC. Mel-frequency cepstral coefficients (MFCCs) are great used as a feature automatic speech recognition. MFCCs feature extraction technique normally includes farming and windowing of signal, applying the Discrete Fourier Transform (DFT) or Fast Fourier Transform (FFT), taking the log of the magnitude, and then compute a Mel-frequency scale on the Mel-filter bank, followed by applying Discrete Cosine Transform (DCT) for log Mel-spectrum into the time domain. The detail description of MFCCs feature extraction various steps are shown in below.

MFCCs feature extraction various steps are:

- First step is farming and windowing. In this step, the analog is segmented into small frames and generate a speech sample of each frame. And then define the windowing for the start and end of each frame.
- Second step is FFT or DFT. In this step, the segmented frame having samples that are converted into the frequency domain (spectrum) and then compute the power spectrum (periodogram) by using Fast-Fourier Transform (FFT).

- Third step is Mel-filter bank. In this step, the power spectrum is changed with the human ear perception of frequency sounds with a linear scale.
- Forth step is DCT. In this step, the Mel-frequency power of the filter bank takes log. And then this step converts the log Mel-spectrum into the time domain. This step generates Mel-frequency cepstral coefficient on time series.

The system applies MFCC as the feature extraction. The aim of applying MFCC over CNN is the performance of MFCC feature extraction is better than CNN in audio recognitions. MFCC feature extraction technique has the several filters and able to extract feature when signal to noise ratio is high. CNN is emphasizing to extract feature from image. CNN is aim to classify short sequences of audio. MFCC can able to extract feature from long audios.

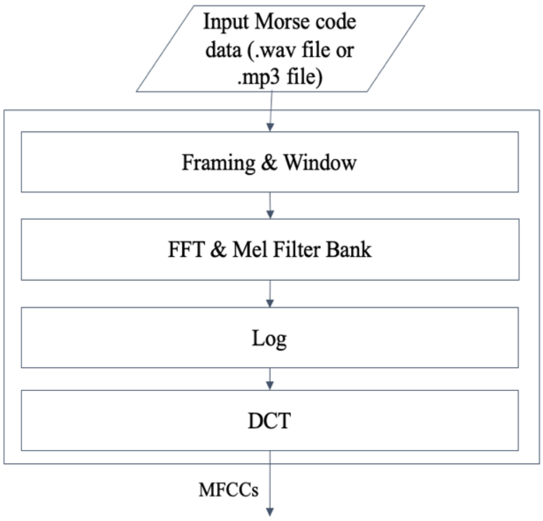


Fig. 2. Simple MFCCs feature extraction structure

## IV. Proposed System Architecture for Morse Code Audio Recognition

The proposed system architecture serves the following processes: dataset preparation, MFCCs feature extraction, Bidirectional LSTM layers, SoftMax, CTC. This system generates the MFCCs features by using the MFCC algorithm for the dataset. Morse code audio dataset was generated by English alphabet text to Morse code audio generator website. The prepared dataset includes training data and testing data. This system normalizes the dataset with features for the LSTM network. LSTM network handles the training process with the incoming dataset. LSTM network generates the related information of each dataset. SoftMax layer makes the classification of the alphabet with LSTM information. Then, CTC decides the related labels with SoftMax data. CTC generates the final result of the proposed recognition system.

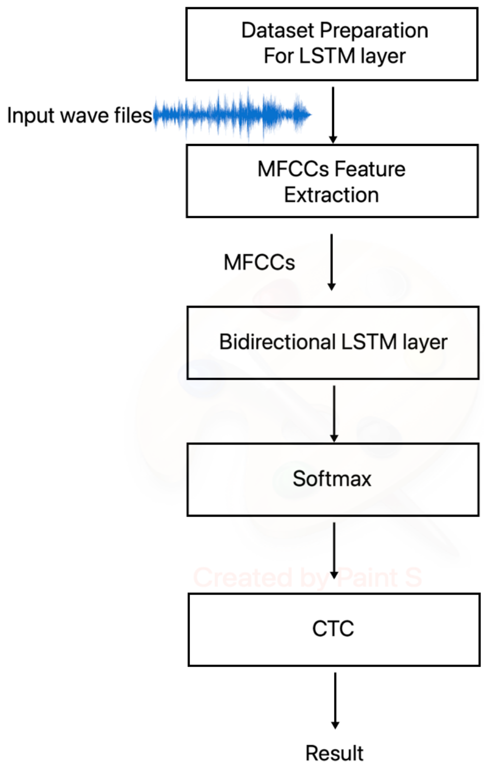


Fig. 3. Proposed system architecture for Morse Code Audio Recognition

### A. Dataset Preparation

Morse code audio dataset consisting of 2300 wave files [12] (training dataset) and 500 wave files [13] (validation and testing dataset). This is generated by English alphabet text to Morse code audio generator website. Every wave file contains a Morse code signal of one or two English alphabet words. The Morse code audio dataset specifications are 16KHz, 16-bit waveform structure, 5 words per minute, and .wav file extension. The labels of each files were stored in the CSV file.

### B. Preprocessing

In the preprocessing process, this system gets the raw Morse code audio wave file to recognize. MFCCs is one of the most generally utilized feature extraction for speech analysis. It is extracted based on the human auditory system, which provides a natural and real reference for speech recognition. This recognition process prepares the dataset as training and testing with labels. The dataset includes training wave files, testing wave files, training labels of each wave file, and testing labels of each wave file. The proposed system of MFCCs feature extraction is shown below.

- Get the frame from wave files.
- The frame has divided the signal into short frames.
- For each frame calculate the periodogram estimate of
- the power spectrum.
- Pass Mel-filter bank to get the power spectra, sum the energy in each filter.
- Take the log of all filter bank of the power spectrum.

- Take the DCT of the log filter bank to the spatial domain.
- Generate MFCCs vector from DCT coefficients.

This system used the python TensorFlow library (MFCC feature extraction algorithm). This system extracts MFCCs features from the dataset by using MFCC feature extraction algorithm. The extracted MFCCs features changes into LSTM network acceptable structure (timesteps, features). Preprocessing contains a splitting of the dataset into input training, target training, and input testing, target testing. Input data are the wave features of audio files and target data are labels of each audio file.

### *C. Morse Code Signal Recognition with LSTM and CTC*

After the preprocessing stage, the system obtains the dataset that contains MFCCs features, information of audio labels. LSTM network cannot accept the vary structure of dataset. Therefore, the standardization is important in this system. The dataset is needed to standardize the standard input structure of the LSTM network. Hence, batch normalization is used to standardize the dataset that is generated by preprocessing stage.

Batch Normalization is utilized in deep learning and take improvement in numerous undertakings. It makes the learning methodology to get a lot of higher learning rates and less careful about initialization. It assists with improving preparing or testing speed and increment the performance.

LSTM layers are added with batch normalization in this system. The difference between LSTM and Bi-LSTM is LSTM can only process the past or current state of information and Bi-LSTM can also process the future information of the current state. Bi-LSTM doesn't need input data to be fixed. Bi-LSTM layers work as training of speech features on timesteps to get related data.

SoftMax layer takes the data to get related alphabet information. CTC layer is used to the prediction of the word label by using the information of the SoftMax layer. CTC result is generated as an English alphabet text.

## V. EXPERIMENTAL RESULT

The paper focus to assist the Morse code listener to get the correct information of the Morse code sender. Morse code transmitter sent Morse code audio and the receiver is difficult to get the right information. The Morse code audio may contain noise and delay. The receiver also needs to learn Morse code. The expert receivers can only know the right information about incoming Morse code audios.

Hence, Morse code audio recognizer model is proposed in this paper. The proposed system model is developed to convert Morse code audio to an English alphabet text format by using MFCC, LSTM, and CTC. The training model is configured by 350 epoch and batch size 10 for training and testing data. The model of the LSTM network got the MFCC features of training data and validation data. Then, the model trained the features inputs to get the target label outputs by using 'tanh' activation function and compiled with CTC loss layer. The model prediction can be predicted by using Keras backend evaluation of CTC decode functions. The prediction results were shown by alphabetic characters.

The two-model results history described as the below Figures. Figure 4 to Figure 7 show the second model result that is configured with epoch 200 and batch size 5. Figure 8 to Figure 11 show the third model result that is configured with epoch 350 and batch size 10. The third model accuracy is better than the second model. The third model of training accuracy is approximately 95 percent in training data. The second model of training accuracy is approximately 60 percent. Figure 4 and Figure 8 show the value loss of model error in training process. Figure 5 and Figure 9 show the value loss of model error in validation process. Figure 6 and Figure 10 show the accuracy of training process. And, Figure 7 and Figure 11 show the accuracy of validation process. The second model of value loss are unstable and high value loss. The third model of value loss are very low in training process and very high in validation process. The second model of accuracy are approximately 60 percent in training process and 7 percent in validation process. The third model of accuracy are approximately 95 percent in training process and 4 percent in validation process. The system is planning to increase validation accuracy by customizing on training data and validation data of training model process. And also, the system is still in progress on calculation of word error rate and calculation the performance of same dataset using similar approaches.

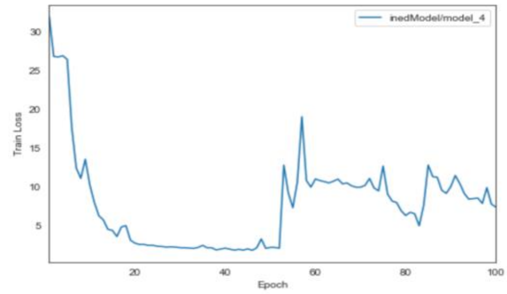


Fig. 4. Training Loss in LSTM model 2 (200 epochs, 5 batch size)

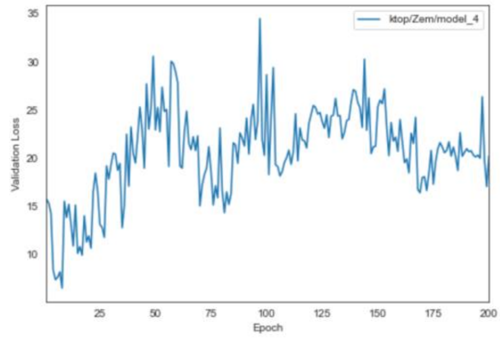


Fig. 5. Validation Loss in LSTM model 2 (200 epochs, 5 batch size)

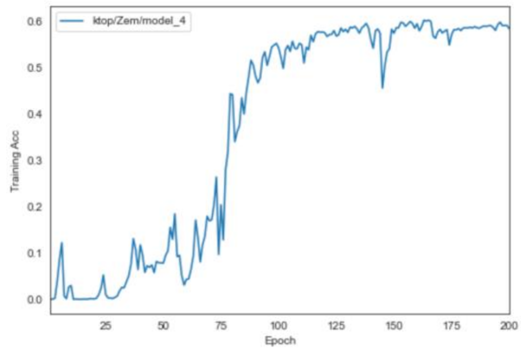


Fig. 6. Training Acc in LSTM model 2 (200 epochs, 5 batch size)

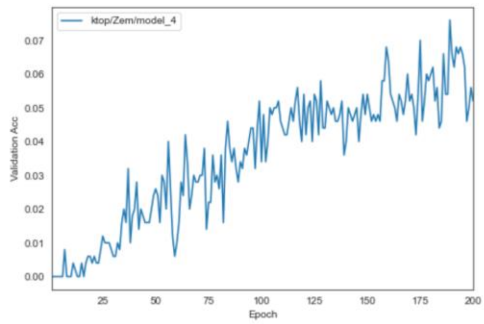


Fig. 7. Validation Acc in LSTM model 2 (200 epochs, 5 batch size)

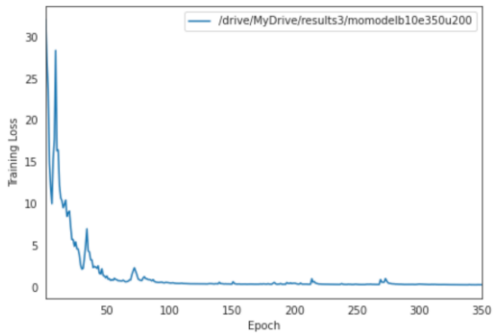


Fig. 8. Training Loss in LSTM model 3 (350 epochs, 10 batch size)

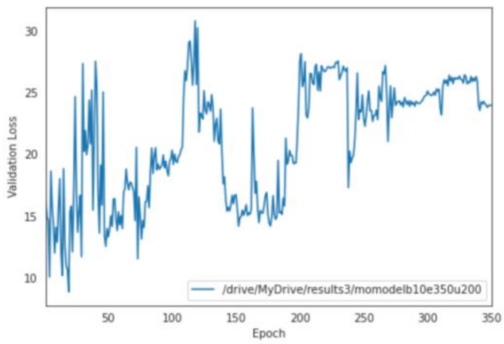


Fig. 9. Validation Loss in LSTM model 3 (350 epochs, 10 batch size)

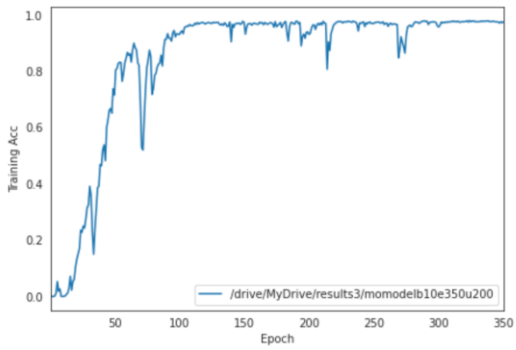


Fig. 10. Training Acc in LSTM model 3 (350 epochs, 10 batch size)

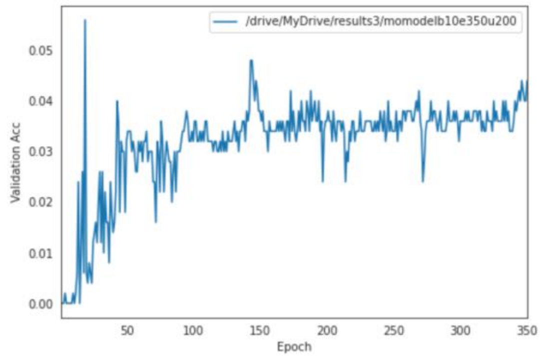


Fig. 11. Validation Acc in LSTM model 3 (350 epochs, 10 batch size)

TABLE I. PERFORMANCE EVALUATION TABLE

| ***Model*** | ***epoch*** | ***Batch size*** | ***Unit*** | ***Traini ng Loss*** | ***Valida tion Loss*** | ***Traini ng Acc*** | ***Validati on Acc*** |
|---|---|---|---|---|---|---|---|
| 1 | 100 | 10 | 100 | 10 | 7.5 | 10% | 7% |
| 2 | 200 | 5 | 100 | 5 | 20 | 60% | 7% |
| 3 | 350 | 10 | 200 | 1 | 25 | 95% | 4% |

## VI. CONCLUSION AND FUTURE WORK

This paper aims to convert from Morse code to human-readable information. (MFCC used) MFCC feature extraction algorithm is used to assist the enhancement of recognition performance. LSTM can be widely used in speech recognition and many other factors. Some of the papers can recognize the tone and space of Morse audio but they cannot recognize with word and alphabets. Some papers used CNN feature extraction instead of MFCCs. Therefore, this paper proposed a deep learning process for Morse code recognition. In the future, the Morse code audio recognition model (MFCC-LSTM-CTC) can assist to recognize on other speech recognition systems. From this experiment, sequence to sequence audio recognition algorithm can solve real-world audio recognition problems.

## REFERENCES


[1] Xianyu Wang, Qi Zhao, Cheng Ma and Jianping Xiong, *"Automatic Morse Code Recognition Under Low SNR",* Advances in Engineering Research, volume 149, 2nd, 2018.

[2] Ming Liu, Yujun Wang, Jin Wang, Jing Wang, Xiang Xie, *"Speech Enhancement Method Based on LSTM Neural Network for Speech Recognition",* School of Information and Electronics, Beijing Institute of Technology, October 2019

[3] De Mauri AG1LE, "Training a Computer to Listen and Decode Morse Code", 2019. [Online]. Available: http://ag1le.blogspot.com/2019/02/training-computer-to-listen-and-decode.html

[4] C-H Yang, L-Y Chuang and T-K Truong, "The application of the neural network on Morse code recognition for users with physical impairments", 2001.

[5] Ericks Rachmat Swedia, Achmad Benny Mutiara, Muhammad Subali, Ernastuti, *"Deep Learning Long-Short Term Memory (LSTM) for Indonesian Speech Digit Recognition using LPC and MFCC Feature",* Departement of Computer Science and Information Technology Gunadarma University and STT Cendekia, October 2018.

[6] Paribesh Regmi, Arjun Dahal, Basanta Joshi, *"Nepali Speech Recognition using RNN-CTC Model",* International Journal of Computer Applications, July 2019.

[7] Aditya Amberkar, Gaurav Deshmukh, Parikshit Awasarmol, Piyush Dave, *"Speech Recognition using Recurrent Neural Network",* MCT's Rajiv Gandhi Institute of Technology, Mumbai, Maharashtra, 2018.

[8] Jason Brownlee, "*LSTM network with python*", machine learning mastery, 2017.

[9] Dong Wang, Xiaodong Wang and Shaohe Lv, *"End-to-End Mandarin Speech Recognition Combining CNN and BLSTM",* National University of Defense Technology, Changsha 410073, China, May 2019.

[10] Sourya Dey, Keith M. Chugg and Peter A. Beerel, "Morse Code Datasets for Machine Learning", Dec 2018

[11] Morse code audio dataset generator website. [Online]. Available: https://morsecode.world/international/translator.html

[12] Training Dataset Google Drive Link. [Online]. Available: https://drive.google.com/drive/folders/10iyq_fDa-r7l1PvAMOJGV27-KiPL-Gx9?usp=sharing

[13] Testing and Validation Dataset Google Drive Link. [Online]. Available: https://drive.google.com/drive/folders/1MJyfFadGmUU_L_ciK9zCgxH91dEIlIMT?usp=sharing

[14] Training csv file Google Drive Link. [Online]. Available: https://drive.google.com/file/d/1vCRheIl4LxGpxKrSHbsbGdsNL1TPMDCn/view?usp=sharing

[15] Testing and Validation csv file Google Drive Link. [Online]. Available: https://drive.google.com/file/d/19LN-gsPsBDCW7tIXvcGTXrPtj7wkHHeT/view?usp=sharing

[16] Understanding architecture of LSTM cell from scratch with code [Online]. Available: https://hackernoon.com/understanding-architecture-of-lstm-cell-from-scratch-with-code-8da40f0b71f4

[17] The frame work of Morse code audio recognition system. [Online]. Available: https://github.com/lucko515/speech-recognition-neural-network